The Mean Value Theorem

> *Rolle's Theorem.* If $f(x)$ is continuous in the closed interval $[a, b]$ and differentiable in the open interval (a, b), and if $f(a) = f(b) = 0$, then there exists a number $c \in (a, b)$ such that $f'(c) = 0$.

When it is represented geometrically, this theorem should strike one as obvious; and to prove it formally may seem a waste of time. Nevertheless, in proving it, we can show how visual logic may be converted into verbal or algebraic logic. Some of the more punctilious issues of Mathematical Analysis cannot be represented adequately in diagrams, and it is for this reason that we prefer to rely upon algebraic methods when seeking firm proofs of analytic propositions.

To prove Rolle's theorem algebraically, we should invoke the result that the function $f(x)$ must achieve an upper bound or a lower bound in the interval (a, b). Or course, the function might have both an upper an a lower bound in (a, b). However, imagine that the function rises above the line in the interval, and that it cuts the line only at the point points a, b which are not included in the open interval. Then it has an upper bound but not a lower bound. If the function is horizontal over the interval, then every point in (a, b) is both an upper bound and a lower bound. Since, in that, case $f'(x) = 0$ for every $x \in (a, b)$ there is nothing to prove.

We shall prove the theorem for the case where there is an upper bound in (a, b) which corresponds to the point c. Then, by assumption, $f(c) \geq f(c + h)$ for small values of $h > 0$ which do not carry us outside the interval. It follows that

$$f'(c+) = \lim_{h \to 0+} \frac{f(c + h) - f(c)}{h} \leq 0.$$

Here the symbolism associated with the limit indicates that h tends to 0 from above.

Now let h be a small *negative* number such that $c + h < c$ remains in the interval (a, b). Then

$$f'(c-) = \lim_{h \to 0-} \frac{f(c + h) - f(c)}{h} \geq 0,$$

where the symbolism associated with the limit indicates that h tends to 0 from below. Now the assumption that $f(x)$ is continuous in (a, b) implies that $f'(c+) = f'(c-)$; and the only way in which this can be reconciled with the inequalities above is if $f'(c) = 0$. This proves the theorem in part. The rest of the proof, which concerns the case where there is a lower bound, follows along the same lines.

The following theorem, which is of prime importance in Mathematical Analysis, represents a generalisation of Rolle's theorem and it has a similar visual or geometric interpretation:

> *The Mean Value Theorem.* If $f(x)$ is continuous in the interval $[a, b]$ and differentiable in the open interval (a, b), then there exists a point $c \in (a, b)$ such that
> $$\frac{f(b) - f(a)}{b - a} = f'(c).$$

Proof. The line passing through the coordinates $\{a, f(a)\}$ and $\{b, f(b)\}$ of the function $f(x)$ has the equation

$$\ell(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a).$$

Define a function $\phi(x) = f(x) - \ell(x)$ which represents the vertical discrepancy between the line and the function.

Since $\ell(x)$ and $f(x)$ agree at the points $x = a, b$, we have $\phi(a) = \phi(b) = 0$. Therefore Rolle's theorem applies to $\phi(x)$ when $x \in (a, b)$—and we may note in passing that ϕ is just the Greek version of f. It follows from the theorem that there exists a point $c \in (a, b)$ at which $\phi'(c) = 0$. That is

$$\begin{aligned}\phi'(c) &= f'(c) - \ell'(c)\\ &= f'(c) - \frac{f(b) - f(a)}{b - a} = 0.\end{aligned}$$

When this equation is rearranged, we have the result which was to be proved.

The mean value theorem can be represented in a way which conforms with some later results. Consider the equation of the linear function $\ell(x)$ which takes the same values as the function $f(x)$ at the points $x = a, b$. Setting $x = b$ in the equation and using $f(b) = \ell(b)$ gives rise to the expression

$$\begin{aligned}f(b) &= f(a) + \frac{f(b) - f(a)}{b - a}(b - a)\\ &= f(a) + f'(c)(b - a),\end{aligned}$$

where c is a value in the interval (a, b), as indicated by the mean value theorem. Let $h = b - a$. Then $c = a + \lambda h$ for some $h \in (0, 1)$, and the equation above can be written in the form of

$$f(a + h) = f(a) + hf'(a + \lambda h).$$

In fact, this expression is valid not only for the points a, b but for any $x, x+h \in [a, b]$; and, therefore, it is appropriate to write

$$f(x+h) = f(x) + hf'(x+\lambda h).$$

Linear Approximations and Newton's Method

The equation above, which provides us with the value of the function f at the point $x+h$, depends upon our knowing the precise point $x+\lambda h$, called the mean value, at which to evaluate the derivative f'. When the derivative is evaluated at the same point x as the function itself is evaluated, we obtain an alternative equation in the form of

$$f(x+h) = f(x) + hf'(x) + r,$$

where r is a remainder term.

This equation is the basis of of Newton's method of approximation which is used for finding the root (ie. the solution) of an algebraic equation $f(x) = 0$. Imagine that ξ is an approximation to the value of x which solves this equation, and let the exact value of the root be denoted by $x = \xi + h$. Then

$$0 = f(\xi + h) = f(\xi) + hf'(\xi) + r.$$

The solution for h is

$$h = -\big\{f'(\xi)\big\}^{-1} f(\xi) - r;$$

and, therefore, the root may be expressed as

$$\xi + h = \xi - \big\{f'(\xi)\big\}^{-1} f(\xi) - r.$$

When seeking to evaluate this expression, we are likely to know all but the final term r. On setting this to zero, we obtain a value ξ_1 on the LHS of the equation which is an approximation to the root, which is liable to be a better approximation than was the original value ξ. To improve the approximation still further, we may replace ξ on the RHS of the equation by ξ_1 so as to obtain ξ_2 on the LHS. The process can be repeated indefinitely in the expectation of generating ever-improving approximations to the root of the equation. In effect, we have specified an iterative method for finding the root. The method is attributable to Newton. The equation of the algorithm by which the $(r+1)$th approximation is found from the rth approximation is

$$\xi_{r+1} = \xi_r - \big\{f'(\xi_r)\big\}^{-1} f(\xi_r).$$

Example. Consider the matter of finding the square root of the number $N > 0$, which is the root of the equation

$$f(x) = x^2 - N = 0.$$

The derivative of the function is $f'(x) = 2x$; and therefore the algorithm above becomes

$$\begin{aligned}\xi_{r+1} &= \xi_r - \frac{1}{2\xi_r}(\xi_r^2 - N)\\ &= \frac{1}{2}\left(\xi_r + \frac{N}{\xi_r}\right).\end{aligned}$$

The starting value, which may denoted by ξ_0, can be an arbitrary positive number, and $\xi_0 = N$ is a reasonable choice. The convergence of the sequence $\{\xi_0, \xi_1, \xi_2, \ldots\}$ to the value of $x = \sqrt{N}$ is remarkably rapid; and, in fact, this algorithm forms part of the library of numerical routines which is built into the ROM of a microcomputer. The following sequence has been generated by a computer which has applied the algorithm to the case where $N = 2$ and $\xi_0 = 2$:

$$\begin{aligned}
\xi_0 &= 2.00000000 & \xi_0^2 &= 4.00000000\\
\xi_1 &= 1.50000000 & \xi_1^2 &= 2.25000000\\
\xi_2 &= 1.41666663 & \xi_2^2 &= 2.00694433\\
\xi_3 &= 1.41421568 & \xi_3^2 &= 2.00000600\\
\xi_4 &= 1.41421354 & \xi_4^2 &= 1.99999993
\end{aligned}$$

After three iterations, the approximation to $\sqrt{2}$ is acceptable. After a few more iterations, one can proceed no further on account of the finite (ie. limited) accuracy with which the computer represents real numbers.

The Quadratic Mean Value Theorem

We have come to regard the mean value theorem as a theorem concerning the approximation of a continuous differentiable function $f(x)$ over the interval $[a, a + h]$ by a linear function $\ell(x)$. Linear approximations are of fundamental importance and are used in many varied contexts. However, they have evident limitations; and, often, it is appropriate to consider more sophisticated approximations which have the potential for being more accurate. Therefore, at least, we should consider using a quadratic function to approximate $f(x)$.

A familiar way of representing a quadratic function is via the equation $p(x) = ax^2 + bx + c$. We shall use other notation. For a start, the letters a, b, c have been preempted for other uses. In the second place, it is helpful to place the "origin" of the argument at the beginning of the interval $[a, b]$. Therefore we shall represent the quadratic by

$$q(x) = q_0 + q_1(x - a) + q_2(x - a)^2.$$

In comparison with the linear function $l(x)$, there is an extra parameter which can be used in pursuit of an improved approximation to $f(x)$ over the interval. We shall continue to apply the endpoint conditions which are that $q(a) = f(a)$ and $q(b) = f(b)$. However, there is a variety of ways in which we may use the additional degree of freedom which is afforded by the extra parameter q_2. The appropriate choice, for present purposes, is to constrain the derivative of $q(x)$ to equal that of $f(x)$ at the point $x = a$. The three conditions which determine the quadratic parameters are therefore

$$\begin{aligned} f(a) &= q(a) = q_0, \\ f'(a) &= q'(a) = q_1, \\ f(b) &= q(b) = q_0 + q_1 h + q_2 h^2, \end{aligned}$$

where $h = b - a$. The parameters q_0 and q_1 are determined immediately by the first two equations, whence the third equation gives

$$q_2 = \frac{f(b) - f(a) - hf'(a)}{h^2}.$$

Thus the approximating quadratic may be written as

$$q(x) = f(a) + (x - a)f'(a) + q_2(x - a)^2,$$

and this becomes

$$q(b) = f(b) = f(a) + hf'(a) + q_2 h^2$$

at the point $x = b$. The result which is known as the quadratic mean value theorem asserts that the parameter q_2 may be expressed in terms of the second-order derivative of the function $f(x)$ evaluated at some point $c \in (a, b)$, which is denoted by $f''(c)$:

> *The Quadratic Mean-Value Theorem.* If $f(x)$ is continuous in the interval $[a, b]$ of length $h = b - a$ and twice differentiable in the open interval (a, b), then there is a point $c \in [a, b]$ such that
>
> $$f(b) = f(a) + hf'(a) + \tfrac{1}{2}h^2 f''(c).$$

Proof. Consider the function $\phi(x) = f(x) - q(x)$ which represents the discrepancy between the function $f(x)$ and its quadratic approximation $q(x)$. Since the two functions agree at the points a, b, it follows that $\phi(a) = \phi(b) = 0$. Therefore Rolle's theorem can be applied to show that there is a point $c_1 \in (a, b)$ such that $\phi'(c_1) = 0$.

Now recall that the parameters of $q(x)$ have been determined to ensure that the condition $f'(a) = q'(a)$ is fulfilled. This implies that $\phi'(a) = 0$. It follow that Rolle's theorem can be applied a second time to the derived function $\phi'(x)$. Thus it transpires that there is a point $c \in (a, c_1)$ at which $\phi''(c) = 0$. That is

$$\begin{aligned} 0 = \phi''(c) &= f''(c) - q''(c) \\ &= f''(c) - 2q_2; \end{aligned}$$

and hence the value $q_2 = \frac{1}{2}f''(c)$ is attributed to the quadratic parameter associated with h^2. ◇

Since the quadratic mean value theorem applies not only in respect of the endpoints of the interval $[a, b]$ but also for any two points $x, x + h \in [a, b]$, it is convenient to use the following expression in representing the result:

$$f(x + h) = f(x) + hf'(x) + \tfrac{1}{2}h^2 f''(x + \lambda h),$$

where $x + \lambda h$, with $\lambda \in [0, 1]$, is some point in the interval $[x, x + h]$.

Taylor's Theorem

The quadratic mean value theorem represents a stepping stone on the way to a general result known as Taylor's theorem. This theorem indicates that, if the function $f(x)$ is n times differentiable over the interval $[a, b]$, then there exists a polynomial $p(x)$ of degree n which agrees with $f(x)$ at the points a and b and which shares with $f(x)$ its derivatives at the point a up to the $(n-1)$th. The nth derivative of this approximating polynomial, which is its final nonzero derivative, can be expressed in term of the nth derivative of $f(x)$ evaluated at some point $c \in (a, d)$.

Taylor's Theorem. If $f(x)$ is a function continuous and n times differentiable in an interval $[a, b]$ of length $h = b - a$, then there exists a point $c \in [a, b]$ such that

$$\begin{aligned} f(b) = f(a) + hf'(a) + \frac{h^2}{2}f''(a) + \cdots \\ \cdots + \frac{h^{(n-1)}}{(n-1)!}f^{(n-1)}(a) + \frac{h^n}{n!}f^n(c). \end{aligned}$$

Proof. Consider approximating $f(x)$ over the interval $[a, b]$ by the polynomial

$$p(x) = p_0 + p_1(x-a) + \cdots + p_{n-1}(x-a)^{n-1} + p_n(x-a)^n.$$

The polynomial may be constrained to agree with $f(x)$ and the endpoints a, b such that $p(a) = f(a)$ and $p(b) = f(b)$. The first of these conditions gives $p_0 = f(a)$. Also, the derivatives of $p(x)$ up to the $(n-1)$th may be constrained to agree with those of $f(x)$ at the point a:

$$p^{(k)}(a) = k!p_k = f^{(k)}(a); \quad k = 1, \ldots, n-1.$$

These equalities provide the coefficients $p_1, \ldots, p_{n-1}$ of the approximating polynomial.

Now define $\phi(x) = f(x) - p(x)$ which represents the error of approximation. Then $\phi(a) = \phi(b) = 0$; as so, by Rolle's theorem, there exists a point $c_1 \in (a, b)$ such that $\phi'(c_1) = 0$. But, by construction, we have $\phi'(a) = 0$, since the first derivatives of $f(x)$ and $p(x)$ have been made to agree at the point a. Therefore Rolle's theorem can be applied again to show that there exists a point $c_2 \in (a, b)$ such that $\phi''(c_2) = 0$. Proceeding in this way, we can show that there is a sequence of points $c_1, c_2, \ldots, c_{n-1}$, with the ordering

$$a < c_{n-1} < \cdots < c_2 < c_1 < b,$$

such that $\phi'(c_1) = \phi''(c_2) = \cdots = \phi^{(n-1)}(c_{n-1}) = 0$. A final application of Rolle's theorem shows that there exists a point $c \in (a, c_{n-1})$ such that

$$\begin{aligned} 0 = \phi^n(c) &= f^n(c) - p^n(c) \\ &= f^n(c) - n!p_n. \end{aligned}$$

This establishes the final coefficient $p_n = f^n(c)/n!$ of the approximating polynomial; and the theorem is proved. $\diamond$

We shall delay using this theorem for a while. Our immediate objective is to exploit the quadratic theorem.

The Maxima and Minima of Twice-Differentiable Functions

Much of economic theory is concerned with problems of optimisation where it is required to find the maximum or the minimum of a twice-differentiable function. In the following discussion, we shall consider only the minimisation of a function of a single variable. Since a problem of maximisation can be solved by minimising the negative of the function in question, there is no real omission in ignoring problems of maximisation. Later we shall consider the optimisation of functions of several variables, including cases where the variables are subject of constraints, ie where the values which can be assigned to the variables are not wholly independent of each other.

We should begin by giving a precise definition of the minimum of a univariate function.

> A point ξ is said to be a *strict minimum* of the function $f(x)$ if $f(\xi) < f(x)$ for all x in an neighbourhood $(\xi - \epsilon, \xi + \epsilon)$ of ξ or, equivalently, if $f(\xi) < f(\xi + h)$ whenever $|h| < \epsilon$ for some small $\epsilon > 0$. The point is said to be a *weak minimum* of $f(x)$ if $f(\xi) \leq f(x)$ for all x in the neighbourhood.

In effect, the point ξ is a strict minimum if the value of f increases with any small departure from ξ, whereas it is a weak minimum if the fails to increase. In general, a function may exhibit these properties at a number of points which are described as *local minima.* If there is a unique point at which the function is lowest, then this is called a *global minimum.*

It is not possible to demonstrate that an analytic function has a global minimum without a complete knowledge of its derivatives of all orders. The conditions which are sufficient for the existence of a local minimum are modest by comparison.

> *Conditions for a Minimum.* A continuous and twice differentiable function $f(x)$ has a minimum at the point ξ if and only if $f'(\xi) = 0$ and $f''(x) \geq 0$ for all x in a neighbourhood $(\xi - \epsilon, \xi + \epsilon)$ of ξ.

Proof. The quadratic mean-value theorem indicates that

$$f(\xi + h) = f(\xi) + hf'(\xi) + \frac{h^2}{2} f''(\xi + \lambda h)$$

for some value $\lambda \in [0, 1]$. Therefore the condition that $f(\xi + h) \geq f(\xi)$ when $|h| < \epsilon$ implies that

$$hf'(\xi) + \frac{h^2}{2} f''(\xi + \lambda h) \geq 0.$$

If $h > 0$, then this implies that

$$f'(\xi) + \frac{h}{2} f''(\xi + \lambda h) \geq 0,$$

and letting $h \to 0+$ shows that $f'(\xi) \geq 0$. On the other hand, if $h < 0$, then dividing by h shows that

$$f'(\xi) + \frac{h}{2} f''(\xi + \lambda h) \leq 0,$$

and letting $h \to 0-$ shows that $f'(\xi) \le 0$. The two inequalities can be reconciled only if $f'(\xi) = 0$.

Now if $f'(\xi) = 0$, then the inequality $f(\xi + h) \ge f(\xi)$ is satisfied for all $|h| < \epsilon$ if and only if $\frac{1}{2}h^2 f''(\xi + \lambda h) \ge 0$ which is if and only if $f''(\xi + \lambda h) \ge 0$. Letting $h \to 0$ establishes that $f''(\xi) \ge 0$ is necessary and sufficient.